International Congress Series 1235 (2002) 73–85

Dynamic nonlinearities in BOLD contrast: neuronal or hemodynamic?

Peter A. Bandettini*, Rasmus M. Birn, Dan Kelley, Ziad S. Saad

Laboratory of Brain and Cognition, Unit on Functional Imaging Methods, National Institute of Mental Health, NIH, Bldg. 10, Rm. 1D80, 10 Center Dr., Bethesda, MD 20892-1366, USA

Abstract

A primary goal of the functional MRI (fMRI) methods development is to characterize the relationship between the blood oxygenation level-dependent (BOLD) signal changes and neuronal activation. Recent studies of blood oxygenation level-dependent (BOLD) signal responses have demonstrated nonlinear behavior with respect to stimulus duration. Specifically, shorter duration stimuli produce larger signal changes than expected from a linear system. The precise reasons for this nonlinearity are not clearly understood. The goal of this study is to further clarify the origin of dynamic BOLD contrast nonlinearities—either neuronal or hemodynamic or both, by a combined approach of task timing modulation, spatial mapping, and modeling/fitting of the BOLD response using the "Balloon model." In this study, we found that (1) in agreement with the literature, the dynamic BOLD "on" response is nonlinear and has significant spatial heterogeneity. Spatial maps of nonlinearity, while highly reproducible, do not correlate with maps of the BOLD response magnitude or latency, but do show some correlation with functional segregation; (2) the dynamic BOLD "off" response is sublinear; and (3) while data fitted with Balloon model hemodynamic parameters, assuming linear neuronal input, generally create nonlinear dynamic BOLD responses, the Balloon model was not able to fit all BOLD contrast response task timing modulations simultaneously. These findings suggest that the dynamic BOLD response may be a linear function of the neuronal input function and that the neuronal input function is not a simple "on/off" boxcar function, but rather a nonlinear function that has an initial overshoot that lasts for approximately 4 s until reaching a steady state.

Keywords: BOLD fMRI; Linearity; Stimulus duration; Visual; Motor; Balloon model

* Corresponding author. Tel.: +1-301-402-1333; fax: +1-301-402-1370.
E-mail address: bandettini@nih.gov (P.A. Bandettini).

0531-5131/02
PII: S0531-5131(02)00174-7

1. Introduction

An increase in neuronal activity leads to a localized increase in the neuronal firing rate, metabolism, blood flow, blood volume, and blood oxygenation, resulting in changes in the local amount of deoxyhemoglobin in each voxel, to which MRI is sensitive. The utility of the functional MRI (fMRI) is directly related to the degree to which neuronal activation can be implied through these hemodynamic changes and the resulting effects on the MRI contrast, most importantly, blood oxygenation level-dependent (BOLD) contrast. An important step in characterizing the relationship between the neuronal firing and measured fMRI signal is by the assessment of the linearity of the measured BOLD signal in response to neural stimulation. While recent studies have suggested that, at steady state (i.e., activation lasting longer than 10 s), the BOLD response shows proportionality to the neuronal firing rate [1] or implied measures of task intensity [2–5], the dynamic BOLD response. Meaning, the magnitude of response as a function of task durations less than 5 s has been shown to have highly nonlinear behavior. The BOLD response does not obey the superposition for certain stimuli [6–8]. While longer duration stimuli behave in an approximately linear fashion, short duration stimuli produce responses larger than the predicted from a linear model.

The purpose of the present work is to examine more closely these nonlinearities and to determine if the source of the nonlinearity is neuronal, hemodynamic, or both. Specifically, we approach this question with three studies. In the first study, we systematically vary the stimulus duration in the motor and visual cortexes and calculate, on a voxel-wise basis, the degree that the response deviates from a linear system. We then compare these "non-linearity" maps with simultaneously derived maps of latency and magnitude (dominated by hemodynamic response variables). If a spatial correlation were found between the linearity maps and the other maps, it would imply that the nonlinearities were of a hemodynamic origin. In the second study, instead of modulating the "on" response, we modulated the "off" response timing and performed a linearity assessment. The assump-

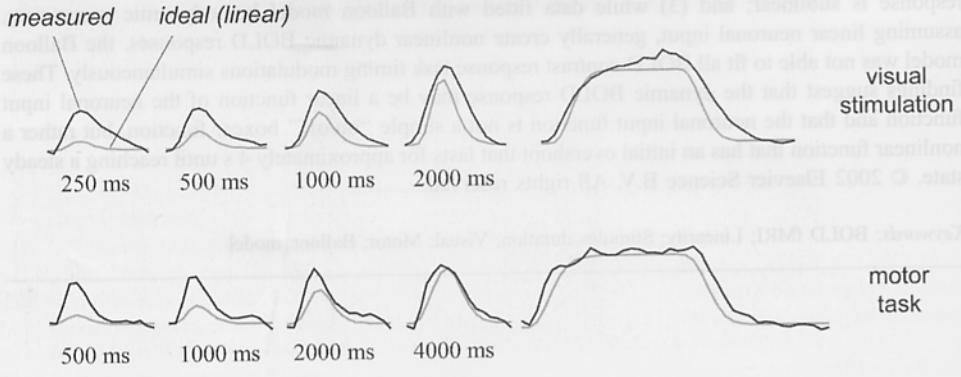


Fig. 1. Top: measured and ideal linear BOLD responses after visual stimulation of 250-, 500-, 1000-, 2000-ms, and 20-s duration. Bottom: measured and ideal responses after finger tapping of 500-, 1000-, 2000-, and 4000-ms, and 20-s duration. Measured and ideal linear responses are also shown superimposed. Short duration stimuli are larger than predicted from a linear system.

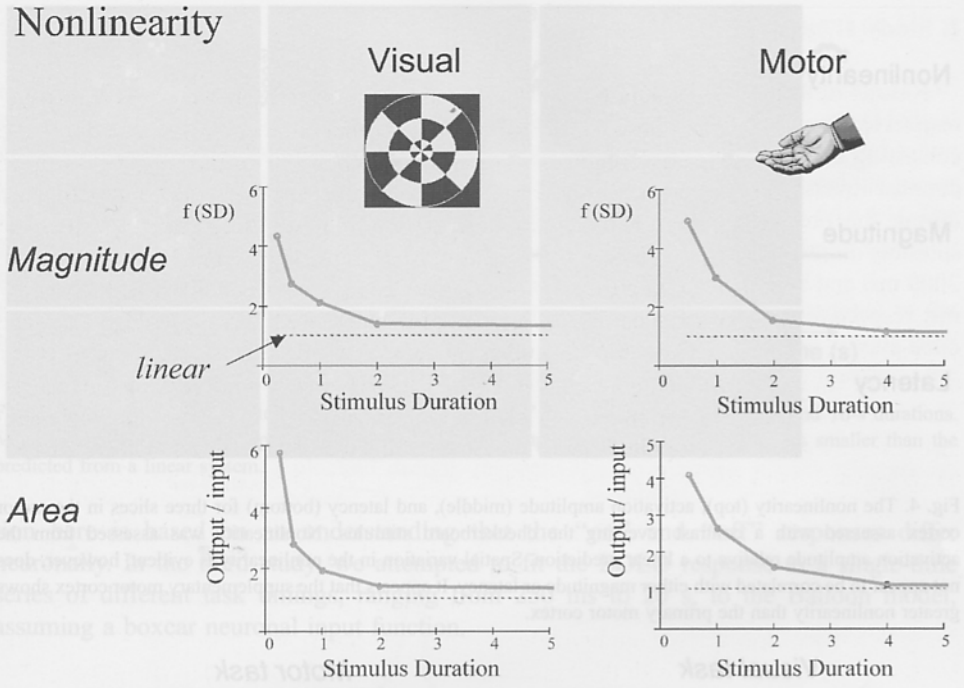


Fig. 2. The amount by which the amplitude (top) and the area (bottom) of the responses are larger at each stimulus duration than the response from a linear system, determined by a linear extrapolation of the responses at the blocked design. In this figure, the nonlinearity curves are averaged over all activated voxels. Left: nonlinearity in the visual cortex; right: nonlinearity in the motor cortex.

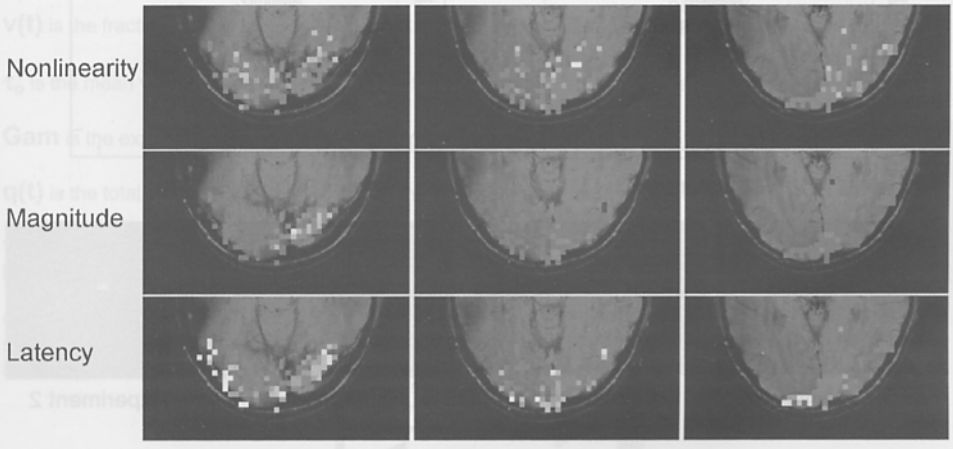


Fig. 3. The nonlinearity (top), activation amplitude (middle), and latency (bottom) for three slices in the visual cortex assessed with a contrast reversing the checkerboard stimulus. Nonlinearity was assessed from the activation amplitude relative to a linear prediction. Spatial variation in the nonlinearity is evident, but does not appear to be correlated with either magnitude or latency.

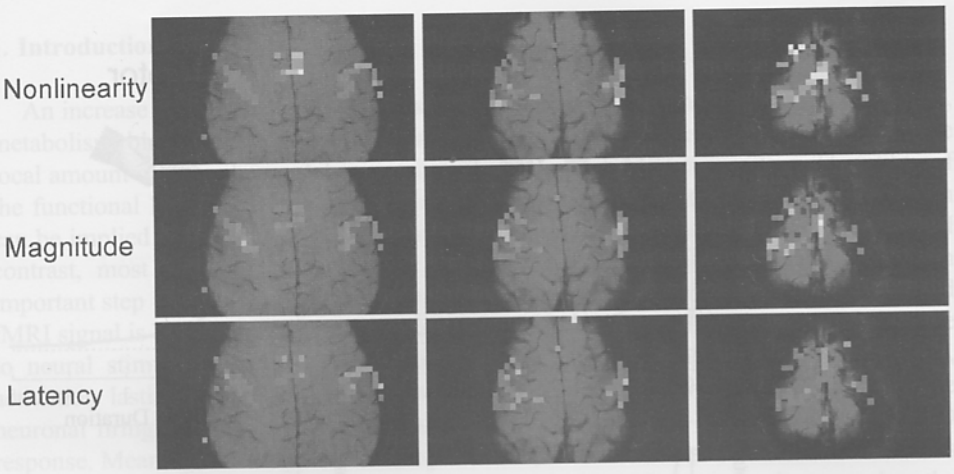


Fig. 4. The nonlinearity (top), activation amplitude (middle), and latency (bottom) for three slices in the motor cortex assessed with a contrast reversing the checkerboard stimulus. Nonlinearity was assessed from the activation amplitude relative to a linear prediction. Spatial variation in the nonlinearity is evident, however, does not appear to be correlated with either magnitude or latency. It appears that the supplementary motor cortex shows greater nonlinearity than the primary motor cortex.

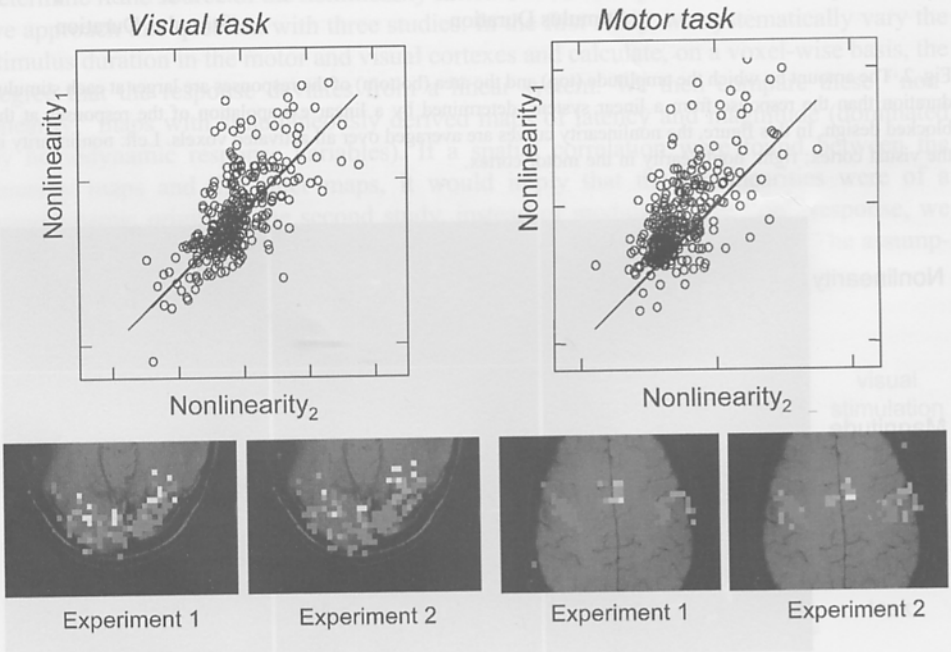


Fig. 5. Computed nonlinearity (as measured by the amplitude of the response compared to a linear model) in two separate runs in the same subject for the visual stimulation experiment (left) and the motor task (right). The line indicates the ideal case of identical nonlinearity values for both runs. The measure of nonlinearity is consistent and reproducible for both tasks. Using the area of the average response instead of the magnitude was equally reproducible.

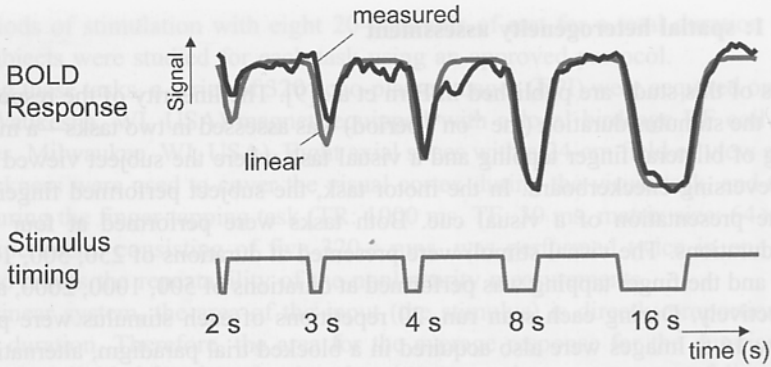


Fig. 6. Measured and ideal linear BOLD responses for "off" durations for 2-, 3-, 4-, 8-, and 16-s durations. Measured and ideal linear responses are also shown superimposed. Short duration "off" is smaller than the predicted from a linear system.

tion here is based on an understanding that the "on" and "off" responses differ neuronally. In the third study, we attempted to fit the BOLD response to a single-time series of different task timings, ranging from 250 ms to 20 s, to the Balloon model, assuming a boxcar neuronal input function.

Balloon Model Parameters

For a given flow of blood into the venous compartment, the three Balloon parameters which control the hemodynamic contribution to the BOLD signal are thought to be:

E_0 represents the fraction of total hemoglobin not bound to O_2;

v(t) is the fraction of voxel volume filled with blood during the active state normalized to that at rest, **V_0**;

τ_o is the mean venous transit time of blood in the venous compartment and equals **V_0 / FlowOut(0)**;

Gam is the exponent defining the relationship between venous outflow and fractional blood volume;

q(t) is the total voxel content of dHB during the active state normalized to that at rest;

viscos is a viscosity term that varies between viscup, during balloon inflation, and viscdown, during balloon deflation.

On a voxelwise basis, the stimulus waveform was smoothed (WAVrisetime), scaled (FLINamp), and phase shifted (FLINdelay) in order to generate an optimally fitting curve, ShiftedFlowIn(t), representing blood flow into the venous compartment.

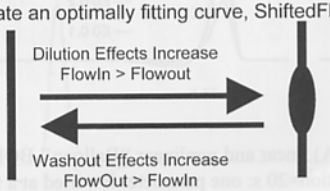


Fig. 7. Balloon model parameters used in the fitting routine in this study.

2. Study 1: spatial heterogeneity assessment

Results of this study are published in Birn et al. [9]. The linearity of the response with respect to the stimulus duration (the "on" period) was assessed in two tasks—a motor task consisting of bilateral finger tapping and a visual task where the subject viewed an 8-Hz contrast reversing checkerboard. In the motor task, the subject performed finger tapping during the presentation of a visual cue. Both tasks were performed at four different stimulus durations. The visual stimuli were presented at durations of 250, 500, 1000, and 2000 ms; and the finger tapping was performed at durations of 500, 1000, 2000, and 4000 ms, respectively. During each scan run, 20 repetitions of each stimulus were presented once every 16 s. Images were also acquired in a blocked trial paradigm, alternating eight

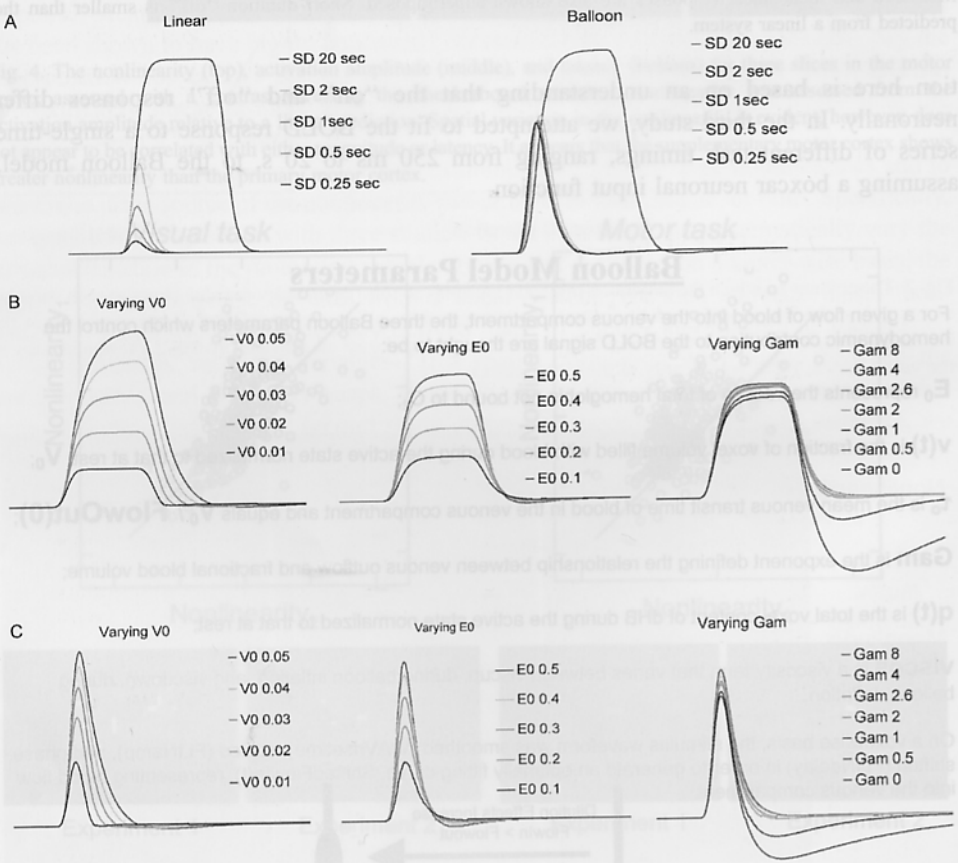


Fig. 8. Simulated BOLD curves: (A) linear and nonlinear "Balloon" BOLD curves for stimulus durations of 20, 2, 1, 0.5, 0.25 s. (B) Stimulus duration=20 s: one parameter is varied at a time. When they are not varied they are set equal to V_0=0.03, E_0=0.3, and Gam=2.6. (C) Stimulus duration=2 s: one parameter is varied at a time. When they are not varied, they are set equal to V_0=0.03, E_0=0.3, and Gam=2.6.

20-s periods of stimulation with eight 20-s periods of rest for a total duration of 320 s. Three subjects were studied for each task using an approved protocol.

During these tasks, a series of 320 echo-planar images (EPI) were acquired on a 3T GE Signa (Waukesha, WI, USA) magnet, equipped with a local birdcage RF coil (Medical Advances, Milwaukee, WI, USA). Eight axial slices with a 24-cm field of view and 5-mm slice thickness were used to cover the visual cortex during the visual task, and the motor cortex during the finger-tapping task (TR: 1000 ms, TE: 30 ms, matrix size: 64×64). The entire experiment, consisting of five 320-s runs, was performed twice in one scanning session to assess the repeatability of the nonlinearity measurements.

In a linear system, the area of the input (the stimulus) is directly proportional to the stimulus duration. Therefore, the area for the average response for the duration of each stimulus was divided by the stimulus duration to produce a measure of linearity—the output of the system for a given level of input. For each voxel, the area of the response as the function of stimulus duration was determined and normalized by the area or amplitude of the fMRI response to a blocked design, respectively. To map the nonlinearity across space, these curves were reduced to one number by computation of the area under the nonlinearity curve. To determine both the magnitude and the latency of the response on a voxel-wise basis, multiple reference functions with varying delays (increments of 100 ms) were generated.

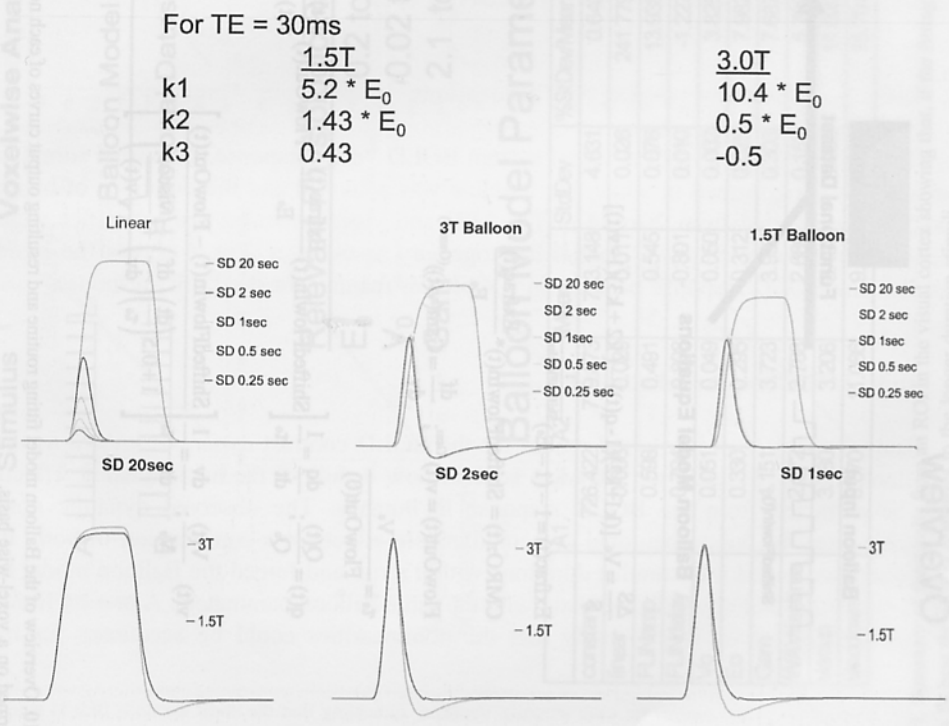


Fig. 9. Balloon model simulations at different field strengths, S.D.=20 s. V_0=0.03, E_0=0.3, and Gam=2.6.

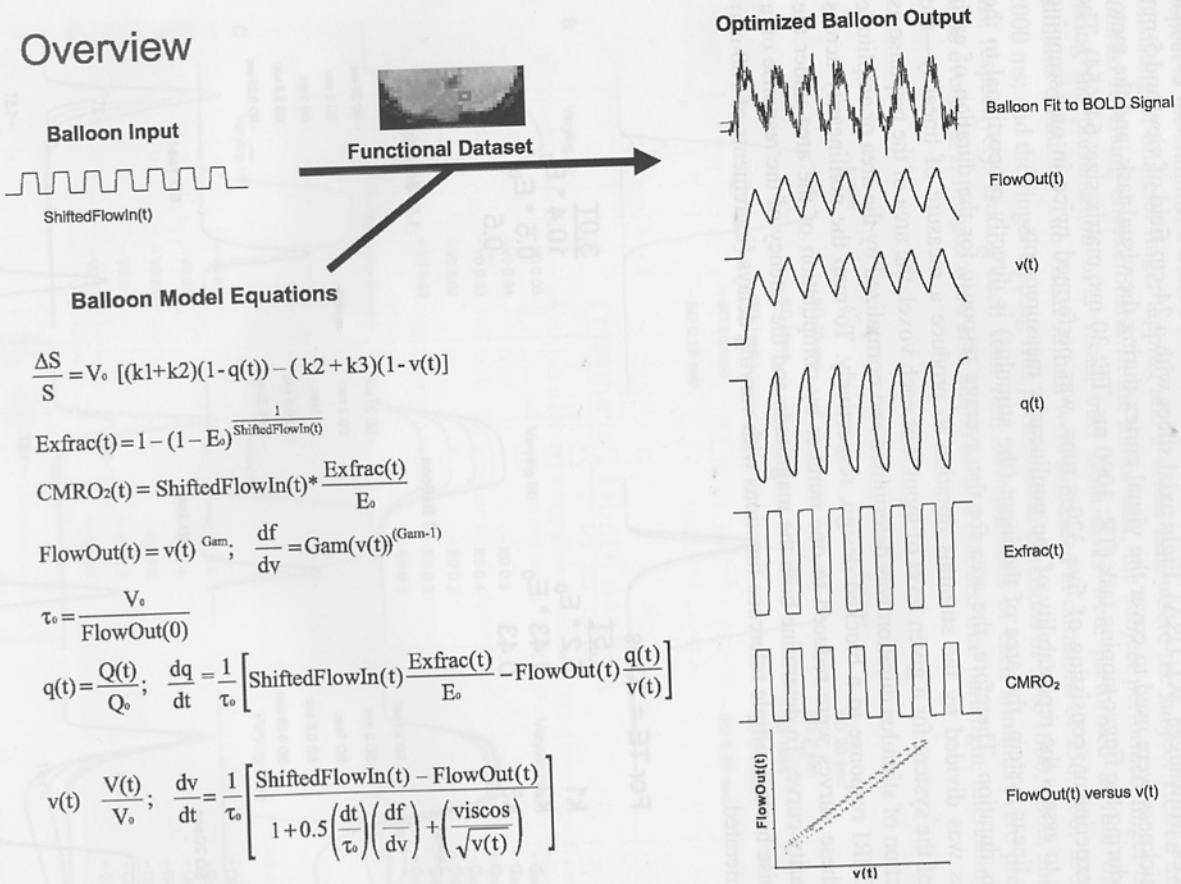
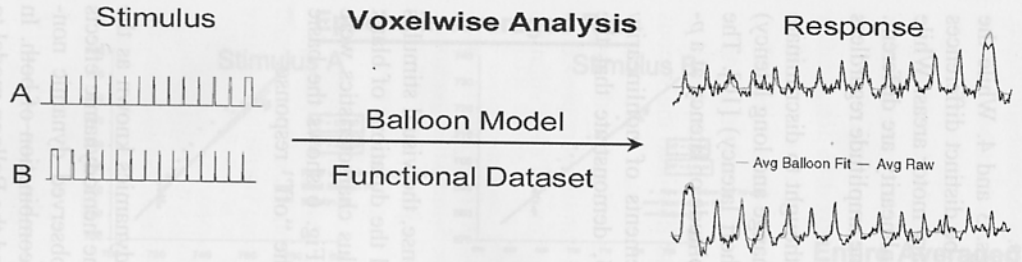


Fig. 10. Overview of the Balloon model fitting routine and resulting output curves of each neurovascular coupling/hemodynamic component that can be obtained. This was performed on a voxel-wise basis.



Relevant Physiologic Range

E_0	0.2 to 0.4
V_0	0.02 to 0.05
Gam	2.1 to 6.4

Balloon Model Parameter Estimation

	A1	A2	Mean	StdDev	%StDev/Mean	B1	B2	Mean	StdDev	%StDev/Mean
constant	726.422	719.873	723.148	4.631	0.640	687.650	695.451	691.551	5.516	0.798
linear	-0.008	0.029	0.011	0.026	241.779	0.023	0.005	0.014	0.013	94.457
FLINamp	0.598	0.491	0.545	0.076	13.938	0.582	0.603	0.592	0.015	2.498
FLINdelay	-0.794	-0.808	-0.801	0.010	-1.227	0.662	0.545	0.604	0.083	13.748
Vo	0.051	0.049	0.050	0.002	3.825	0.034	0.041	0.037	0.004	12.007
Eo	0.330	0.295	0.312	0.025	7.982	0.436	0.393	0.415	0.030	7.288
Gam	4.151	3.723	3.937	0.303	7.687	3.742	3.495	3.618	0.175	4.830
WAVrisetime	2.572	2.788	2.680	0.153	5.706	2.431	2.625	2.528	0.138	5.445
viscup	3.780	3.206	3.493	0.406	11.620	8.529	7.115	7.822	1.000	12.782
viscdown	8.870	11.086	9.978	1.567	15.704	9.945	10.250	10.098	0.215	2.133

Fig. 11. Summary of the results from an ROI in the visual cortex showing that, if the fitting parameters are allowed to vary within a physiologically reasonable range, the curves fit best the BOLD responses to the longer duration stimuli.

In agreement with previous studies, the BOLD response was found to be nonlinear for stimuli under a 5-s duration, with activation amplitudes larger than the predicted from a linear model at shorter stimulus durations. This is illustrated in Fig. 1, which shows the BOLD response averaged over all stimulation epochs and activated voxels for stimulus duration and the responses predicted from a linear system. The amount by which the area and the magnitude of the responses at stimulus duration are larger than expected from a linear model is shown in Fig. 2 for both the visual and motor tasks.

Maps of nonlinearity, latency, and magnitude are shown in Figs. 3 and 4. While the degree of nonlinearity is spatially heterogeneous, there appears to be distinct differences between the degree of nonlinearity in the primary and supplementary motor areas. While responses in both areas are nonlinear, the manifestations of the nonlinearity are different. The responses in the supplementary motor cortex are almost the same amplitude regardless of the stimulus duration.

The response latency and percentage signal change have been thought to discriminate between signals from large veins (with a large percentage signal change and long latency) and small vessels (with a small percentage signal change and short latency) [10]. The nonlinearity is not significantly correlated with either response amplitude or latency at a p-value of 0.01.

The high correlation between repeated voxel-wise measurements of nonlinearity (significant at a p-value of 0.001), which are shown in Fig. 5, demonstrate that the variability of the nonlinearity is not an artifact of noise.

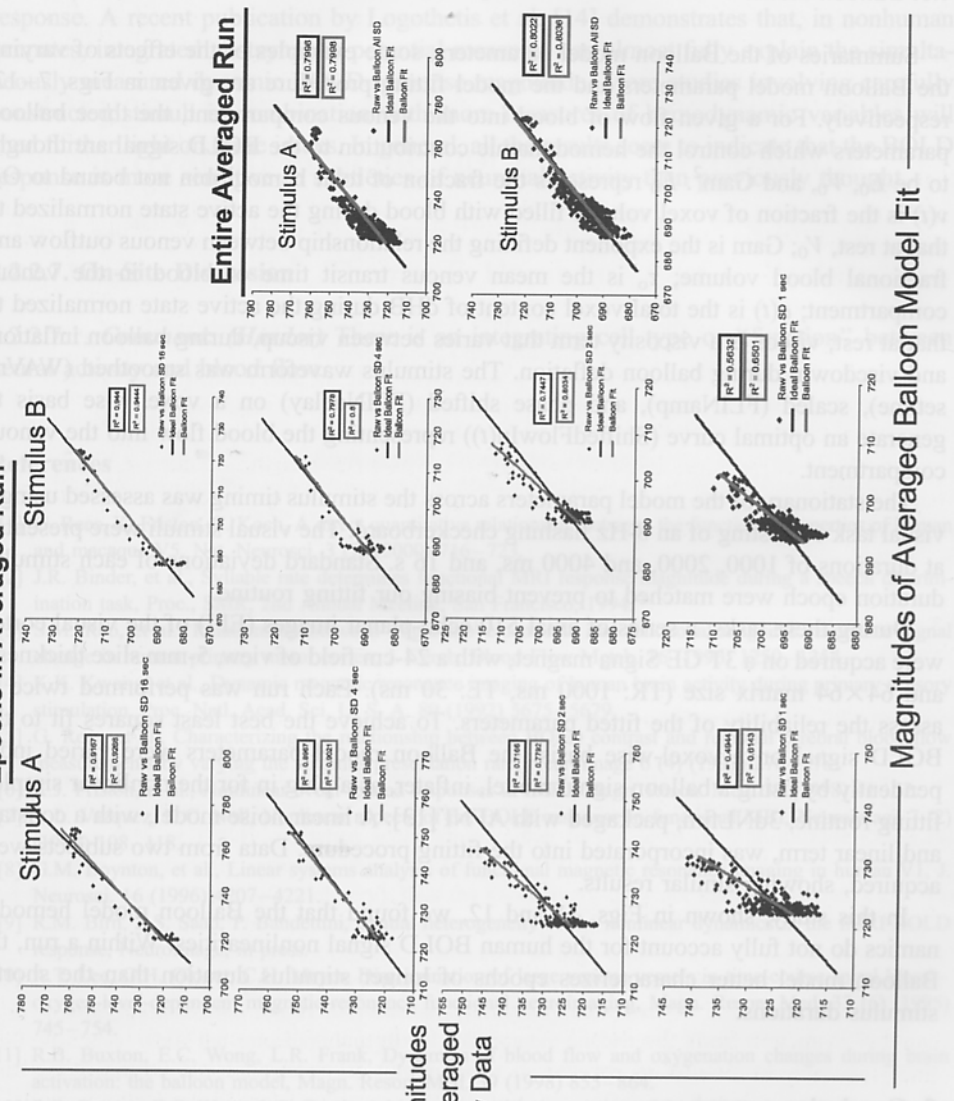
3. Study 2: response to varying "off" durations

To assess the linearity of the visual system BOLD "off" response, the visual stimulus paradigm was performed in which the baseline was activated and the durations of blank fixation were varied from 2 to 16 s. The MRI and visual stimulus characteristics were identical to those in Study 1: spatial heterogeneity assessment. Fig. 6 shows the basic results. The BOLD response behaves in a sublinear manner for the "off" response.

4. Study 3: Balloon model studies

Using a physiologically based model for the BOLD contrast dynamics known as the "Balloon model" [11,12], we intend to address how details of the hemodynamic effects can be extracted from the BOLD response in humans. The observed dynamic nonlinearities may be due to neuronal or hemodynamic effects or a combination of both. In this study, we varied the stimulus durations within a run and forced the Balloon model to fit all stimulus durations within a voxel with the same balloon parameters. A precise fit for all stimulus durations would imply that the nonlinearities could be accounted for by

Fig. 12. Breakdown of the results for each stimulus duration, indicating that the short duration BOLD signal changes are consistently underestimated by the Balloon model fit when the same parameters are applied to all durations simultaneously.


Epochs in Averaged Run
Stimulus A
Stimulus B
Entire Averaged Run
Stimulus A
Stimulus B
Magnitudes of Averaged Raw Data
Magnitudes of Averaged Balloon Model Fit
Raw vs Balloon SD 16 sec
Raw vs Balloon SD 4 sec
Raw vs Balloon SD 2 sec
Raw vs Balloon SD 1 sec
Raw vs Balloon All SD
Ideal Balloon Fit
Balloon Fit

hemodynamic factors alone. A breakdown of the fit would imply that the neuronal input function for the model (a simple on–off boxcar function) is inaccurate.

4.1. Balloon model

Summaries of the Balloon model parameters, some examples of the effects of varying the Balloon model parameters and the model fitting procedure are given in Figs. 7–12, respectively. For a given flow of blood into the venous compartment, the three balloon parameters which control the hemodynamic contribution to the BOLD signal are thought to be E_0, V_0, and Gam^2. E_0 represents the fraction of total hemoglobin not bound to O_2; $v(t)$ is the fraction of voxel volume filled with blood during the active state normalized to that at rest, V_0; Gam is the exponent defining the relationship between venous outflow and fractional blood volume; τ_o is the mean venous transit time of blood in the venous compartment; $q(t)$ is the total voxel content of dHB during the active state normalized to that at rest; viscos is a viscosity term that varies between viscup, during balloon inflation, and viscdown, during balloon deflation. The stimulus waveform was smoothed (WAVrisetime), scaled (FLINamp), and phase shifted (FLINdelay) on a voxel-wise basis to generate an optimal curve (ShiftedFlowIn(t)) representing the blood flow into the venous compartment.

The stationary of the model parameters across the stimulus timing was assessed using a visual task consisting of an 8-Hz flashing checkerboard. The visual stimuli were presented at durations of 1000, 2000, and 4000 ms, and 16 s. Standard deviations of each stimulus duration epoch were matched to prevent biasing our fitting routine.

During these tasks, a series of axial 510 echo-planar images (EPI) of the visual cortex were acquired on a 3T GE Signa magnet, with a 24-cm field of view, 5-mm slice thickness, and 64×64 matrix size (TR: 1000 ms, TE: 30 ms). Each run was performed twice to assess the reliability of the fitted parameters. To achieve the best least squares fit to the BOLD signal on a voxel-wise basis, the Balloon model parameters were varied independently by using a balloon signal model, inflater, as a plug in for the nonlinear simplex fitting routine, 3dNLfim, packaged with AFNI [13]. A linear noise model, with a constant and linear term, was incorporated into the fitting procedure. Data from two subjects were acquired, showing similar results.

In this study, shown in Figs. 11 and 12, we found that the Balloon model hemodynamics do not fully account for the human BOLD signal nonlinearities. Within a run, the Balloon model better characterizes epochs of longer stimulus duration than the shorter stimulus durations.

5. Conclusion

For brief stimulus "on" periods, signal increases are larger than expected. These nonlinearities show considerable yet reproducible spatial heterogeneity that does not correlate with the hemodynamic latency or magnitude maps. For brief stimulus "off" periods, signal decreases are smaller than expected from a linear system. We also found that the Balloon model hemodynamics do not fully account for the human BOLD signal

dynamic nonlinearities. Within a run for a given stimulus, the Balloon model is better at characterizing epochs of longer stimulus duration than shorter stimulus duration.

In general, these studies imply that the BOLD response nonlinearities may be explained by a combination of a nonlinear neuronal input and nonlinearities in the hemodynamic response. A recent publication by Logothetis et al. [14] demonstrates that, in nonhuman primates, integrated postsynaptic potential measures can almost fully explain the simultaneously measured dynamic BOLD response magnitude. More studies involving carefully constructed stimuli in combination with more measures of hemodynamic variables will shed further light on these issues. In general, all the results seem to indicate that the BOLD response is more sensitive to subtleties of neuronal activity than previously thought.

2.2.2.7. On-Site Discussion

2.2.2.7.1. ***Comment: (Harder)*** There is an integrating cell type or "function" between neural activity and blood flow.

References

[1] G. Rees, K. Friston, C. Koch, A direct quantitative relationship between the functional properties of human and macaque V5, Nat. Neurosci. 3 (7) (2000) 716–723.
[2] J.R. Binder, et al., Syllable rate determines functional MRI response magnitude during a speech discrimination task, Proc., SMR, 2nd Annual Meeting, San Francisco, 1994.
[3] S.M. Rao, et al., Relationship between finger movement rate and functional magnetic resonance signal change in human primary motor cortex, J. Cereb. Blood Flow Metab. 16 (1996) 1250–1254.
[4] K.K. Kwong, et al., Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 5675–5679.
[5] G. Rees, et al., Characterizing the relationship between BOLD contrast and regional cerebral blood flow measurements by varying the stimulus presentation rate, Neuroimage 6 (4) (1997) 270–278.
[6] K.J. Friston, et al., Nonlinear event-related responses in fMRI, Magn. Reson. Med. 39 (1998) 41–52.
[7] A.L. Vazquez, D.C. Noll, Nonlinear aspects of the BOLD response in functional MRI, Neuroimage 7 (2) (1998) 108–118.
[8] G.M. Boynton, et al., Linear systems analysis of functional magnetic resonance imaging in human V1, J. Neurosci. 16 (1996) 4207–4221.
[9] R.M. Birn, Z.S. Saad, P. Bandettini, Spatial heterogeneity of the nonlinear dynamics in the fMRI BOLD response, Neuroimage, in press.
[10] A.T. Lee, G.H. Glover, C.H. Meyer, Discrimination of large venous vessels in time-course spiral blood–oxygen-level-dependent magnetic-resonance functional neuroimaging, Magn. Reson. Med. 33 (6) (1995) 745–754.
[11] R.B. Buxton, E.C. Wong, L.R. Frank, Dynamics of blood flow and oxygenation changes during brain activation: the balloon model, Magn. Reson. Med. 39 (1998) 855–864.
[12] R.B. Buxton, E.C. Wong, L.R. Frank, A biomechanical interpretation of the BOLD signal time course: the balloon model, Proc., ISMRM 5th Annual Meeting, Vancouver, 1997.
[13] R.W. Cox, AFNI: software for analysis and visualization of functional magnetic resonance neuroimages, Comput. Biomed. Res. 29 (3) (1996) 162–173.
[14] N. Logothetis, et al., Neurophysiological investigation of the basis of the fMRI signal, Nature 412 (2001) 150–157.